

**AMENDMENTS TO THE CLAIMS**

This listing of claims will replace all prior versions, and listings, of claims in the application:

1. (Currently Amended) A method, comprising:  
receiving a request on a DNS server from a client for a web page at a first web address, the first web address including a hostname;  
determining traffic loads of a plurality of mirrored customer web servers each addressable by the requested hostname among a customer's plurality of web servers, each of the customer web servers storing the web page;  
determining a customer web server from the plurality of mirrored customer web servers that is appropriate for the request, the customer web server having a traffic load that is optimal over lower than traffic loads of remaining customer web servers from the plurality of mirrored customer web servers according to optimal customer web server selection rules defined on a per domain name basis;  
determining an IP address of the customer web server;  
sending the IP address of the customer web server to the client;  
receiving a request from the client for static content on the web page at a second web address, the second web address specifying a network of caching servers;  
determining service metrics of a set of caching servers each addressable by the second web address in the network of caching servers, the network of caching servers does not include the customer's plurality of web servers;  
wherein a customer is a customer of a service for use of the network of caching servers managed by the service that stores static content for the customer;  
determining a caching server from the set of caching servers that is appropriate for the request for static content, the caching server having service metrics that are optimal over better than service metrics of remaining caching servers from the set of caching servers according to optimal caching server selection rules defined on a per domain name basis;  
determining an IP address of the caching server; and  
delivering the IP address of the caching server to the client.
2. (Previously Presented) The method of claim 1 further comprising:

determining load of caching servers in the network of caching servers;  
wherein the determining the caching server from the network of caching servers that is  
appropriate for the request step selects a caching server having a latency and a load lower  
than latency or load of remaining caching servers from the network of caching servers.

3. (Previously Presented) The method of claim 1 further comprising:  
determining whether the caching server includes the static content;  
determining a customer web server that includes the static content when the caching server does  
not include the static content;  
retrieving the static content from the customer web server that includes the static content; and  
storing the static content from the customer web server in the caching server.

4. (Previously Presented) The method of claim 3 wherein the determining the customer web  
server that includes the static content step comprises:  
determining traffic loads of the plurality of mirrored customer web servers, each of the customer  
web servers storing the static content; and  
determining a second customer web server from the plurality of mirrored customer web servers  
that is appropriate for the request, the second customer web server having a traffic load  
lower than traffic loads of remaining customer web servers from the plurality of mirrored  
customer web servers.

5. (Previously Presented) The method of claim 4 wherein retrieving the static content from  
the customer web server step comprises:  
determining an IP address of the second customer web server; and  
requesting the static content from the second customer web server at the second customer web  
server IP address.

6. (Previously Presented) The method of claim 1 wherein the network of caching servers includes a domain name server.

7. (Previously Presented) The method of claim 1 wherein the request from the client for the web page is transferred from a first domain name server;  
wherein the network of caching servers includes a second domain name server; and  
wherein the second domain name server determines the customer web server from the plurality of mirrored customer web servers.

8. (Currently Amended) A method, comprising:  
receiving a first request on a DNS server from a client DNS server to resolve a first domain name, the client DNS server receiving a request from a client of a web page address that includes the first domain name;  
determining load measurements of a plurality of mirrored customer web servers each addressable by the first domain name among a customer's plurality of web servers, each of the customer web servers addressable by the first domain name, and each of the customer web servers configured to service the request from the client;  
determining a customer web server from the plurality of mirrored customer web servers, the customer web server having a traffic load that is optimal over ~~lower than~~ traffic loads of other customer web servers from the plurality of mirrored customer web servers according to optimal customer web server selection rules defined on a per domain name basis;  
determining an IP address of the customer web server;  
providing the IP address of the customer web server to the client DNS server;  
receiving a second request from the client DNS server to resolve a second domain name, the client DNS server receiving a request from the client of a uniform resource locator obtained from the web page associated with the web page address that includes the second domain name;

determining performance metric measurement of a set of caching servers each addressable by the second domain name in a network of caching servers, the network of caching servers does not include the customer's plurality of web servers;

wherein a customer is a customer of a service for use of the network of caching servers managed by the service that store static content for the customer;

determining a caching server from the set of caching servers, the caching server having performance metrics that are optimal over ~~lower than~~ performance metrics of other caching servers from the set of caching servers according to optimal caching server selection rules defined on a per domain name basis; and

delivering an IP address of the caching server to the client DNS server.

9. (Original) The method of claim 8 wherein the load measurements comprise latency measurements.
10. (Previously Presented) The method of claim 8 wherein the performance metric measurements comprise any of: load CPU and memory measurements, HTTP response measurements, and FTP response measurements.
11. (Previously Presented) The method of claim 8 further comprising:  
in response to receiving the uniform resource locator request at the caching server, determining whether the caching server includes the data;  
retrieving data from a second customer web server from the mirrored customer web servers when the caching server does not include the data; and  
storing the data within the caching server.

12. (Previously Presented) The method of claim 11 wherein retrieving data from the second customer web server step comprises:

determining the second customer web server from the plurality of mirrored customer web servers, the second customer web server having a traffic load lower than traffic loads of remaining customer web servers from the plurality of mirrored customer web servers; and retrieving the data from the second customer web server.

13. (Previously Presented) The method of claim 8 further comprising:

receiving a first request from a second client DNS server to resolve a third domain name, the second client DNS server receiving a request from a second client of a second web page address that includes the third domain name;

determining load measurements of a plurality of second mirrored customer web servers each addressable by the third domain name among a customer's plurality of web servers, and each of the second customer web servers storing data configured to service the request from the second client;

determining a second customer web server from the plurality of second mirrored customer web servers, the second customer web server having a traffic load lower than traffic loads of other second customer web servers from the plurality of second mirrored customer web servers;

determining an IP address of the second customer web server; and

providing the IP address of the second customer web server to the second client DNS server.

14. (Previously Presented) The method of claim 13 further comprising:

receiving a second request from the second client DNS server to resolve the second domain name, the second client DNS server receiving a request from the second client of a second uniform resource locator that includes the second domain name;

determining performance metric measurement of the set of caching servers;

determining a second caching server from the set of caching servers, the second caching server having performance metrics lower than performance metrics of other caching servers from the set of caching servers; and  
delivering an IP address of the second caching server to the second client DNS server.

15-20. (Canceled)

21. (Currently Amended) An apparatus, comprising:

a DNS server that receives a request from a client for a web page at a first web address, the first web address including a hostname;

a traffic load subsystem that determines traffic loads of a plurality of mirrored customer web servers each addressable by the requested hostname among a customer's plurality of web servers, each of the customer web servers storing the web page;

a customer web server selection subsystem that determines a customer web server from the plurality of mirrored customer web servers that is appropriate for the request, the customer web server having a traffic load that is optimal over ~~lower than~~ traffic loads of remaining customer web servers from the plurality of mirrored customer web servers according to optimal customer web server selection rules defined on a per domain name basis;

an IP address subsystem that determines an IP address of the customer web server;

a request directing subsystem that directs the request from the client to the customer web server;

a request receiving subsystem that receives a request from the client for static content on the web page at a second web address, the second web address including the hostname;

a service metric subsystem that determines service metrics of a set of caching servers each addressable by the requested hostname in a network of caching servers, the network of caching servers does not include the customer's plurality of web servers;

wherein a customer is a customer of a service for use of the network of caching servers managed by the service that store static content for the customer;

a caching server selection subsystem that determines a caching server from the set of caching servers that is appropriate for the request for static content, the caching server having service metrics that are optimal over ~~better than~~ service metrics of remaining caching

servers from the network of caching servers according to optimal caching server selection rules defined on a per domain name basis;

a caching server IP address determining subsystem that determines an IP address of the caching server; and

a caching server IP address delivery subsystem that delivers the IP address of the caching server to the client.

22. (Previously Presented) The apparatus of claim 21 further comprising:

a caching server load determination subsystem that determines load of caching servers in the set of caching servers;

wherein the caching server load determination subsystem selects a caching server having a latency and a load lower than latency or load of remaining caching servers from the set of caching servers.

23. (Previously Presented) The apparatus of claim 21 further comprising:

a caching server content determination subsystem that determines whether the caching server includes the static content;

a customer web server content determination subsystem that determines a customer web server that includes the static content when the caching server does not include the static content;

a customer web server content retrieval subsystem that retrieves the static content from the customer web server that includes the static content; and

a caching server content storage subsystem that stores the static content from the customer web server in the caching server.

24. (Previously Presented) The apparatus of claim 23 wherein the customer web server content determination subsystem comprises:

a customer web server traffic load determination subsystem that determines traffic loads of the plurality of mirrored customer web servers, each of the customer web servers storing the static content; and

a secondary customer web server selection subsystem that determines a second customer web server from the plurality of mirrored customer web servers that is appropriate for the request, the second customer web server having a traffic load lower than traffic loads of remaining customer web servers from the plurality of mirrored customer web servers.

25. (Previously Presented) The apparatus of claim 24 wherein the customer web server content retrieval subsystem comprises:

a secondary customer web server IP address determination subsystem that determines an IP address of the second customer web server; and

a secondary server content requesting subsystem that requests the static content from the second customer web server at the second customer web server IP address.

26. (Previously Presented) The apparatus of claim 21 wherein the network of caching servers includes a domain name server.

27. (Previously Presented) The apparatus of claim 21

wherein the client for the web page is transferred from a first domain name server;

wherein the network of caching servers includes a second domain name server; and

wherein the second domain name server determines the customer web server from the plurality of mirrored customer web servers.

28. (Currently Amended) An apparatus, comprising:

- a DNS server that receives a first request from a client DNS server to resolve a first domain name, the client DNS server receiving a request from a client of a web page address that includes the first domain name;
- a traffic load subsystem that determines load measurements of a plurality of mirrored customer web servers each addressable by the first domain name among a customer's plurality of web servers, and each of the customer web servers configured to service the request from the client;
- a customer web server selection subsystem that determines a customer web server from the plurality of mirrored customer web servers, the customer web server having a traffic load that is optimal over ~~lower than~~ traffic loads of other customer web servers from the plurality of mirrored customer web servers according to optimal customer web server selection rules defined on a per domain name basis;
- an IP address subsystem that determines an IP address of the customer web server;
- wherein the IP address subsystem sends the IP address of the customer web server to the client DNS server;
- wherein the DNS server receives a second request from the client DNS server to resolve a second domain name, the client DNS server receiving a request from the client of a uniform resource locator obtained from the web page associated with the web page address that includes the second domain name;
- a caching server performance metric subsystem that determines performance metric measurement of a set of caching servers each addressable by the second domain name in a network of caching servers, the network of caching servers does not include the customer's plurality of web servers;
- wherein a customer is a customer of a service for use of the network of caching servers managed by the service that store static content for the customer;
- a caching server selection subsystem that determines a caching server from the set of caching servers, the caching server having performance metrics that is optimal over ~~lower than~~ performance metrics of other caching servers from the set of caching servers according to optimal caching server selection rules defined on a per domain name basis; and
- a caching server selection subsystem that delivers an IP address of the caching server to the client DNS server.

29. (Previously Presented) The apparatus of claim 28 wherein the load measurements comprise latency measurements.
30. (Previously Presented) The apparatus of claim 28 wherein the performance metric measurements comprise any of: load CPU and memory measurements, HTTP response measurements, and FTP response measurements.
31. (Previously Presented) The apparatus of claim 28 further comprising:  
a caching server data determination subsystem that, in response to receiving the uniform resource locator request at the caching server, determines whether the caching server includes the data;  
a second customer web server data retrieval subsystem that retrieves data from a second customer web server from the plurality of mirrored customer web servers when the caching server does not include the data; and  
a caching server data storing subsystem that stores the data within the caching server.
32. (Previously Presented) The apparatus of claim 31 wherein the second customer web server data retrieval subsystem comprises:  
a second customer web server determining subsystem that determines the second customer web server from the plurality of mirrored customer web servers, the second customer web server having a traffic load lower than traffic loads of remaining customer web servers from the plurality of mirrored customer web servers; and  
a second customer web server data retrieval subsystem that retrieves the data from the second customer web server.

33. (Previously Presented) The apparatus of claim 28:

wherein the DNS server receives a first request from a second client DNS server to resolve a third domain name, the second client DNS server receiving a request from a second client of a second web page address that includes the third domain name;

wherein the traffic load subsystem determines load measurements of a plurality of second mirrored customer web servers each addressable by the third domain name among a customer's plurality of web servers, and each of the second customer web servers storing data configured to service the request from the second client;

wherein the customer web server selection subsystem determines a second customer web server from the plurality of second mirrored customer web servers, the second customer web server having a traffic load lower than traffic loads of other second customer web servers from the plurality of second mirrored customer web servers;

wherein the IP address subsystem determines an IP address of the second customer web server; and

wherein the IP address subsystem sends the IP address of the second customer web server to the second client DNS server.

34. (Previously Presented) The apparatus of claim 33 wherein the DNS server receives a second request from the second client DNS server to resolve the second domain name, the second client DNS server receiving a request from the second client of a second uniform resource locator that includes the second domain name;

wherein the caching server performance metric subsystem determines performance metric measurement of the set of caching servers;

wherein the caching server selection subsystem determines a second caching server from the set of caching servers, the second caching server having performance metrics lower than performance metrics of other caching servers from the set of caching servers; and

wherein the caching server selection subsystem delivers an IP address of the second caching server to the second client DNS server.